R-513-NASA

April 1971

A STUDY OF USER BEHAVIOR IN PROBLEMSOLVING WITH AN INTERACTIVE COMPUTER

M. J. Seven, B. W. Boehm and R. A. Watson

A Report prepared for

NATIONAL AERONAUTICS AND SPACE ADMINISTRATION

Rand
SANTA MONICA, CA. 90406

This research is sponsored by the National Aeronautics and Space Administration under Contract No. NAS-12-2144. Views or conclusions contained in this study should not be interpreted as representing the official opinion or policy of NASA or of The Rand Corporation.

R-513-NASA

April 1971

A STUDY OF USER BEHAVIOR IN PROBLEMSOLVING WITH AN INTERACTIVE COMPUTER

M. J. Seven, B. W. Boehm and R. A. Watson

A Report prepared for

NATIONAL AERONAUTICS AND SPACE ADMINISTRATION

Rand
SANTA MONICA, CA. 90406

Rand maintains a number of special subject bibliographies containing abstracts of Rand publications in fields of wide current interest. The following bibliographies are available upon request:

Africa • Arms Control • Civil Defense • Combinatorics
Communication Satellites • Communication Systems • Communist China
Computing Technology • Decisionmaking • Delphi • East-West Trade
Education • Foreign Aid • Foreign Policy Issues • Game Theory
Health-related Research • Latin America • Linguistics • Maintenance
Mathematical Modeling of Physiological Processes • Middle East
Policy Sciences • Pollution • Program Budgeting
SIMSCRIPT and Its Applications • Southeast Asia • Systems Analysis
Television • Transportation • Urban Problems • USSR/East Europe
Water Resources • Weapon Systems Acquisition
Weather Forecasting and Control

To obtain copies of these bibliographies, and to receive information on how to obtain copies of individual publications, write to: Communications Department, Rand, 1700 Main Street, Santa Monica, California 90406.

Published by The Rand Corporation

PREFACE

This Report presents the salient results of a controlled experiment in man-computer problemsolving, and discusses their connections to related experiments and to general problems in evaluating computer system performance. This analysis was one of a group of computer system analysis and simulation studies sponsored by the National Aeronautics and Space Administration; a summary of the overall research effort is given in

B. W. Boehm, *Computer Systems Analysis Methodology: Studies in Measuring, Evaluating, and Simulating Computer Systems*, The Rand Corporation, R-520-NASA, September 1970.

Detailed results of the other component studies are reported in

N. R. Nielsen, *ECSS: An Extendable Computer System Simulator*, The Rand Corporation, RM-6132-NASA, February 1970.

R. A. Watson, *Computer Performance Analysis: Applications of Accounting Data*, The Rand Corporation, R-573-NASA/PR, April 1971.

T. E. Bell, *Computer Performance Analysis: Measurement Objectives and Tools*, The Rand Corporation, R-584-NASA/PR, February 1971.

This Report should be of interest to researchers and computer system designers and managers concerned with problems of man-computer interaction.

SUMMARY

An exploratory investigation tested the effects of forced temporal lockout intervals on user performance in an interactive man-computer problemsolving situation. Twenty subjects performed a planning task using the JOSS* interactive computer system as a decisionmaking aid. In the lockout conditions, the subject's terminal was mechanically locked out of the system for a specified length of time after each trial, i.e., after he had received a current set of results.

In general, the subjects having a 5-min lockout period after each trial not only achieved better solutions to the problem than did the control (no lockout) group, but they also used far less computer time and less total working time. A longer lockout period (8 min) appeared to be disruptive, especially for more experienced subjects.

Other findings suggest that self-imposed restraint, such as that resulting from a restrictive charge algorithm, can also improve problemsolving efficiency, and that the users' acceptance of the system is not necessarily a valid predictor of system effectiveness.

---

*JOSS is the Trademark and Service Mark of The Rand Corporation.

CONTENTS

FIGURES

TABLES

I. INTRODUCTION

During the next few years, the cost of computer time will further decrease, more nonprogrammers will demand direct access to computers, and computers will be used increasingly to augment human problemsolving and decision processes. Therefore, future systems must be designed to utilize the user's time more efficiently.

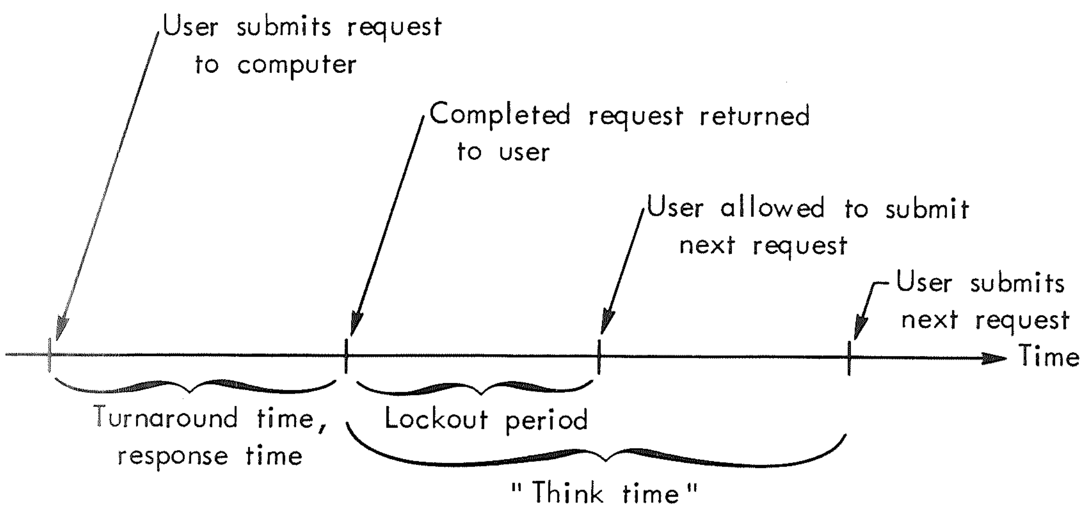
Although many systems have been designed expressly to augment human thought processes, the designs are based less on data than on some mixture of art and tradition. Sackman [1,2] and Gold [3] have reviewed a small group of pioneering studies that sought to add experimental evidence to the philosophical arguments on both sides of the controversy involving the relative effectiveness of time-sharing and batch processing. Although the results of these studies are less than conclusive and at times in disagreement with one another, they clearly demonstrate that (a) individual differences are important--indeed, they can overshadow differences in computer systems; (b) the problem being solved interacts with the system used to solve it; and (c) little is known about the parameters of significance in the man-machine-problem interface.

THE PRESENT STUDY

Our investigation of problemsolving behavior with an interactive computer system is one of a group of computer system analysis and simulation studies sponsored by NASA. The problem selected for study was not a programming problem; rather, it was a problem similar to that faced by a controller or planner using an interactive computer to assist his decisionmaking. The primary objective was to gain some insight into the problems of structuring more extensive experiments to study the relevant parameters of the man-machine-problem interface. Since some experimental framework was required for the exploration, the study was built around an experimental investigation of several notions expressed or implied in existing literature. The primary treatment to be studied was forced lockout from the computer system; however, the study also provided limited data relating to variable versus fixed time intervals, acceptability versus effectiveness, and the effect of charging algorithms.

Forced Lockout

One of the most interesting challenges to the implicit "common sense" assumptions regarding desirable system response characteristics stemmed from the work of Gold [4]. After finding that his subjects' performances improved more *between* sessions at the console than *during* such sessions, he hypothesized that a forced temporal spacing or "lockout" between computer messages and human responses might stimulate more creative thinking and hence more effective problemsolving. The diagram below defines the lockout period and its relationship to other common terms in interactive computing; imposing a lockout period on a user requires him to spend a certain amount of "think time." Presumably, however, beyond some optimum lockout interval, further delay would disrupt the user's thoughts and waste his time.



Definition of terms in man-computer interaction

Fixed versus Variable Delays

Carbonell [5] noted that uncertainty regarding the cost (in terms of user time) of a given computer run affects the ability of the user to make good operational decisions. He cited an unpublished observation by D. G. Bobrow that, in a time-sharing environment, users prefer a fixed delay to a possibly shorter but variable one. Carbonell, *et al.* [6] contended that unpredictable conditions disturb the user and interfere with his efficient use of the computer.

Acceptability and Effectiveness

It is commonly assumed in the literature that the greater the acceptability of the system to the user, the more effectively he will use it. Such a contention is supported in part by findings that show such a correlation (e.g., Sackman and Gold [7]); however, the evaluations may be inconclusive because the systems were evaluated *after* the results had been obtained, and "winners" are apt to be very tolerant. Carbonell, *et al.* [6] tend to equate accomplishment with satisfaction, but they do maintain some reservation by suggesting an empirical check of the degree of correspondence.

Effect of Charging Algorithms

Nickerson, *et al.* [8] suggest that a user's strategy for interacting with a system will depend on the way in which system charges are made (e.g., on-line versus central processor time). The contention appears reasonable, both logically and empirically, judging from informal observations and personal experience. However, the authors then venture the prescription that the charging algorithm *should* be such that it is in the user's best interest to use the system much as he would if it were free. Again, the implication is that the consumer's desire is a reliable guide to optimum system design.

II. METHOD

The general problemsolving situation required the subject to solve a geographical area servicing problem with the aid of JOSS,* Rand's interactive computer system. Subjects were allowed 2 hr to solve an open-ended problem for which a range of solutions existed. A protocol of each subject's performance was generated from automatic recordings within the JOSS system, records handwritten by an observer, and audio tape recordings of all the subject's vocalizations. The resulting data were examined using analysis of variance and regression techniques.

THE TEST PROBLEM

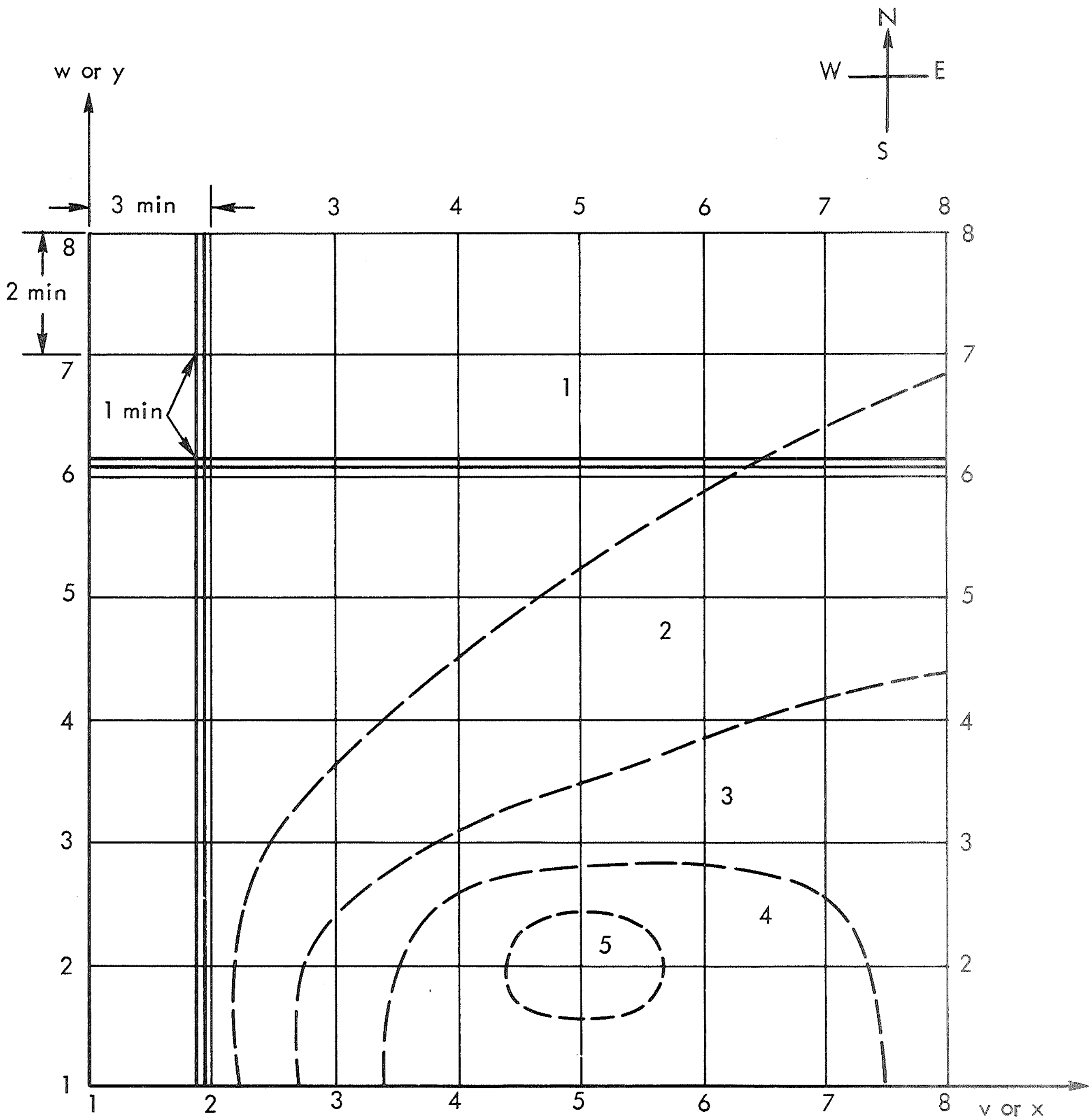
Each subject was given a map showing a grid of surface streets, two freeways, and some contour lines that indicated the frequencies of emergencies per day per intersection throughout the area (Fig. 1). Transit times between intersections were defined to be 2 min on north-south surface streets, 3 min on east-west surface streets, and 1 min on the freeways. It was possible to enter or leave the freeways at any intersection, but a time penalty of 1 min was assessed for each transition.

The subject's task was to designate three surface intersections at which to locate three emergency hospitals and to specify a set of decision rules stipulating when and when not to use the freeways in dispatching ambulances from these hospitals. His goal was to minimize the average response time per emergency for the entire area, taking into account the different accident densities. His solution was subject to the constraint that the maximum one-way response time to any given location be no more than 12 min. It was made clear that the number of ambulances was unlimited; scheduling and turnaround time were not factors in the problem.

The JOSS system was preprogrammed to print, on demand, an evaluation of the effectiveness of location and decision rule inputs. Other data relating to the problem were also available. Hospital locations were to be designated in X, Y coordinates shown on the map. Variables to be used

---

*JOSS is the Trademark and Service Mark of The Rand Corporation.



Transit times between intersections
1 min on freeway
2 min on north-south surface streets
3 min on east-west surface street
Time to both enter (1 min) and leave (1 min) freeway system from surface streets: 2 min

Fig. 1 — Problem map

in the decision rules* were specified so that the subject could refer to particular hospitals (i = 1, 2, or 3), hospital locations (x, y), or emergency locations (v, w) in terms understood by the special program. If so instructed ("Do part 1."), the program performed an evaluation computation that provided the average response time per emergency and the maximum response time to any emergency. If requested to do so, the program also provided various types of information matrices: (a) one matrix showing minimum response time to each intersection from any of the three hospitals ("Do part 210."), (b) three individual matrices showing response times to each intersection from each of the three hospitals ("Do part 220."), and (c) an individual matrix showing response times from any hospital specified ("Do part 221 for i = __."). The subjects were given an annotated reproduction of an actual JOSS record of three "trials" (Fig. 2) to familiarize them with the special program.

In addition to the special program, the normal JOSS capabilities were available to the subject. Since the study was not construed as a test of the subject's skill in the JOSS language, an observer provided guidance, as needed, on matters specific to the system (e.g., logical forms). A list of forbidden variables prevented the subject from inadvertently using variables that had been used in the control program.

TREATMENT GROUPS

The primary experimental treatment was provided by programming the JOSS system to lock the subject out of the system for a specified length of time after each trial, i.e., after a "calculate" cycle that provided an assessment of a set of inputs. During this interval, the keyboard was locked and the red signal lamp was lighted. At the end of the interval, the "ready" beep sounded, the green signal lamp lighted, and control of the keyboard was returned to the subject. Lockout conditions were different for each of five groups of subjects, and included both fixed and variable intervals. Based upon preliminary work with the test problem and the hypotheses being tested, the following treatment groups were specified.

---

*A variable (z) was to be set at 1 if the freeway was to be used, or at 0 if it was not to be used.

Example: Using the JOSS program

X(1)=3
Y(1)=3
X(2)=6
Y(2)=3
X(3)=5
Y(3)=6

Hospital locations

100.1 Set z=1 if v=2 or w=6. — a decision rule
200 Do part 210. — a print option
Do part 1.

You supply these

Average = 5.40496
Maximum = 16

16	7	10	7	4	7	10	13
14	6	8	5	2	5	8	11
6	5	4	3	2	3	4	5
10	6	4	5	2	4	7	10
8	6	2	5	4	2	5	8
6	5	0	3	3	0	3	6
8	6	2	5	5	2	5	8
10	7	4	7	7	4	7	10

Shortest response times to intersections

Program types out these

Y(3)=7
200 Do part 221 for i=3.
Do part 1.

You supply revisions

Average = 5.44628
Maximum = 14

14	9	8	5	2	5	8	11
12	8	6	3	0	3	6	9
8	7	6	5	4	5	6	7
16	8	10	7	4	7	10	13
18	9	12	9	6	9	12	15
20	10	14	11	8	11	14	17
22	11	16	13	10	13	16	19
24	12	18	15	12	15	18	21

Response times from hospital 3

Program responds

Y(3)=8
100.2 Set z=1 if w=5.
200 Do part 210.
Do part 1.

You supply revisions

Average = 5.89256
Maximum = 14

12	10	6	3	0	3	6	9
14	9	8	5	2	5	8	11
9	8	8	7	6	7	8	9
10	7	10	9	8	9	10	11
8	6	2	5	5	2	5	8
6	5	0	3	3	0	3	6
8	6	2	5	5	2	5	8
10	7	4	7	7	4	7	10

Shortest response times

Program responds

Fig. 2—Sample JOSS printout

Group 0

Group 0 received no lockout or other constraint and used the computer system freely. This group served as the control group for assessing the effects of the treatments used with the other four groups. The way these subjects used the system could be considered the "natural" approach.

Group C

The "Choice" group received no forced lockout, but each member was instructed to minimize the amount of time he used the system (both console time and computer time). The subject could log out of the system (or in again) at will by typing the words "Type timer". It was intended that this group would show the effects of a self-imposed constraint of the type that might result from a restrictive charging algorithm.

Group 5

Group 5 was locked out of the computer system for 5 min following each trial. The subjects knew when the lockouts would occur and how long they would last. The 5-min time interval was intended to be mildly disruptive.

Group 8

Group 8 was locked out of the computer system for 8 min following each trial. Again, the subjects knew when the lockouts would occur and how long they would last. The 8-min time interval was intended to be more severely disruptive.

Group V

Group V was locked out of the system for variable lengths of time from 0 to 10 min, with successive pairs of trials averaging 5 min. The subjects knew when the lockouts would occur and the range of possible durations, but had no foreknowledge of any particular lockout duration.

Although these subjects would have, on the average, a lockout period approximately equal to that of Group 5, the period was variable and uncertain. (The actual sequence of lockout times for these subjects was 8, 2, 5, 5, 2, 8, 4, 6, 8, 2, 4, and 6 min.)

SUBJECTS

The 20 subjects were Rand employees representing a variety of scientific and technical disciplines--primarily graduate students employed for the summer, but also some full-time staff members with several years of work experience. Programming experience ranged from none to 8 yr; experience with interactive systems, none to 4 yr. Formal training in operations analysis ranged from none to 15 courses. The subjects were ranked according to an estimate of relevant capability and experience, based on questionnaire responses and supervisor ratings. The experience measure was then used to divide the subjects into five balanced groups of four subjects each. The five groups were then randomly assigned to the lockout conditions.

As a result of having to replace six subjects who were dropped for a variety of reasons,* the experience balance was only approximate. The final composition of the groups is shown in Table 1. The aggregate experience of Group 5 was the lowest (larger numbers indicating less experience) and that of Group C was the highest.

APPARATUS AND MATERIALS

Subjects were tested two at a time in nonadjacent offices. In addition to the normal office equipment, the offices were furnished with a JOSS typewriter terminal, audio recording equipment, and various problem materials.

---

*Two subjects were dropped because of oversights in the original procedures or instructions (one "dumped" the control program), two were lost because JOSS shut down during their runs, one was unable to work with an observer in the room, and one preferred to work without computer assistance.

Table 1

DISTRIBUTION OF EXPERIENCE RANKS BY GROUPS

Experience Ranks	Group 0	Group C	Group 5	Group V	Group 8
Ranks[a]	1	3	5	2	4
	10	8	9	7	6
	11	12	14	13	15
	20	16	17	18	19
Sum	42	39	45	40	44

[a]Rank 1 = Most experience.

JOSS Terminal

The JOSS terminal consists of a mobile console holding a standard IBM Selectric typewriter (with a slightly modified character set) and an auxiliary control box. The unit is connected to the central computer system via jacks located in the offices of staff members.

Audio Recording Equipment

During each problem session, the subject wore a lavaliere microphone to record his vocalizations continuously. The observer's comments were recorded similarly.

Problem Materials

At the beginning of the problem run, the subject received a six-page information package that included a brief discussion of the background of the problem, an explicit statement of the problem to be solved, instructions on how to use the special JOSS program, an example of several interactions with the special program (Fig. 2), and an annotated street map that defined the problem situation. The subject had access to as many copies of the street map as he desired. In addition, the subjects were given scratch paper, pens, pencils, and marking pens in several colors. Other materials* were available on request.

*It had been anticipated that subjects might wish to construct overlays; however, none requested such materials.

Questionnaires

The pretest questionnaire obtained information regarding the subject's experience with computers, programming, and operations analysis. Questions were asked and responses recorded by the experimenters. The post-test questionnaire, completed by the subject, gathered information on his views of the test situation and on his method of solving the problem.

Observer's Logs

During the problem run, the observer kept a running time log of the subject's actions and comments. The information was recorded on 8.5-by-11-in. forms on which problem time, in quarter minutes, was preprinted along the left margin. Log entries included such items as "types x, y's," "writes on scratch paper," and "draws boundary on printout."

PROCEDURE

Prior to Formal Test

The test procedures and program were developed in pilot runs with the experimenters serving as subjects. Then the subjects for the formal experiment were recruited, given a brief explanation of the experiment, and requested to familiarize themselves with the JOSS system. Their experience rankings were determined and the balanced groups formed. Assignment of lockout conditions, testing times, test rooms, and observers was done using randomization techniques.

During the Test

At their scheduled times, the subjects met with members of the study team for a briefing on the problem. The objectives of the study were outlined, the data collection methods described, and the subjects assured that no information regarding their individual performances would be released to anyone. The subjects were encouraged to "think out loud" while working on the problem, if they could do so comfortably.

The performance goal was stated indefinitely; the subjects could quit any time they were satisfied with their results during the 2-hr session. To provide some frame of reference, all subjects were told that an average response time of 4.67 min was the best time achieved to date, but that better ones might exist.* (The time given as a target was equivalent to a criterion score of 90 on a 100-point scale.)

The subjects were then taken to the test rooms and the test begun. The observer signed onto the JOSS system, readied the files, and started the timer subroutine. The subject received the problem package and began reading. The observer turned on the tape recorder and then took a position where he could see what was going on without intruding on the subject's work area. Throughout the session, the observer maintained the activity log and answered allowable questions. The subject was told when 1 hr had passed and when 10 min remained; when time** was up, he was asked to stop work, and the debriefing session was begun.

The debriefing session consisted of the subject's completing the post-test questionnaire and providing any other comments or information he thought relevant. The subject was cautioned not to discuss with anyone the problem or any aspects of his work until the test schedule was completed. He was thanked for his participation and told he would receive a summary of the results when they were compiled. Although the session was officially ended, many subjects resumed work on the problem for their own satisfaction.

MEASURES

The principal dependent measures were quality of best solution (minimum average response time), elapsed user time to best solution, total user time expended, and total machine time expended. However, to serve the exploratory objectives of the study, many other measures relating to the subjects' problemsolving behavior were collected. Most of

---

*The minimum time was actually 4.52 min; the best obtained by one of the subjects was 4.57 min.

**In some cases, the 120-min time limit was extended as much as 10 min to compensate for time lost due to interruptions or malfunctions.

these measures were obtained directly from the JOSS printouts, some from analysis of the observer's logs, some from the questionnaires, and some from combining or transforming other measures.

III. RESULTS AND DISCUSSION

Before presenting the results, a cautionary note is in order. Considering the exploratory nature of the study, the semiformal test environment, the restricted sample, and the possible uniqueness of this particular problem or the JOSS system, the results must be considered tentative.

EXPERIMENTAL ANALYSES

The analyses presented in this section primarily concern the factors related to the evaluation of the effects of the experimentally controlled variables. In keeping with the exploratory nature of the study, a significance level of 10 percent or less was considered an adequate indicator of an effect worth considering further.

Quality of Solution

Figure 3 shows the distribution of scores on the criterion measure (minimum average response time) for the 20 subjects grouped according to lockout treatments. The mean score for each group is also indicated. The number next to each data point is the experience ranking of that particular subject.

Perhaps the most striking aspect of Fig. 3 is that the variance of scores differs widely from group to group. This difference is significant at the .05 level ($F_{max} = 73$). The forced lockout may have contributed to this reduction of dispersion due to individual differences.

Figure 3 also indicates an apparent difference in group means due to lockout conditions and an apparent effect due to experience. To test the significance of the lockout and experience factors, the data were subjected to an analysis of variance using five levels of lockout and two levels of experience in a two-way analysis with two replications per cell.* The two higher-ranked subjects were grouped into a

*See Ref. 9, p. 87.

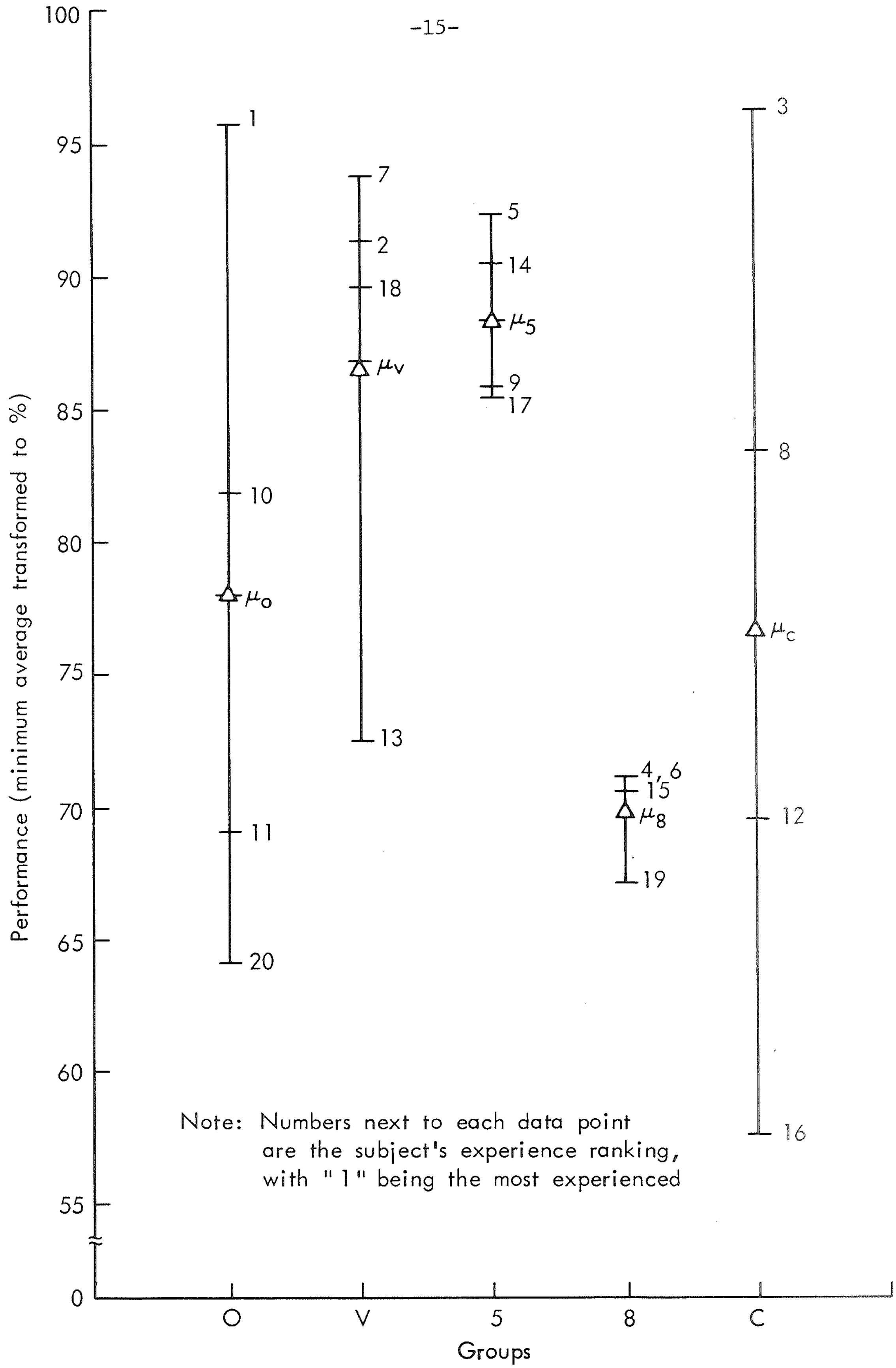


Fig. 3—Criterion scores attained by the subjects

"higher experience" subset and the two lower-ranked subjects into a "lower experience" subset within each group. The results of this analysis appear in Table 2.

Table 2

ANALYSIS OF VARIANCE SUMMARY: CRITERION SCORES

Source	Sum of Squares	Degrees of Freedom	Mean Square	F	Significance Level
Total	.282	19	--	--	--
Lockout	.096	4	.024	5.11	.025
Experience	.082	1	.082	17.45	.005
L × E	.057	4	.014	2.98	.100
Error	.047	10	.0047	--	--

Lockout is significant at the .025 level, experience is significant at the .005 level, and the interaction between lockout and experience is significant at the .10 level. In view of the heterogeneity of variance indicated by F_{max}, these results must be interpreted with caution; however, the robustness of the analysis of the variance model and the uncertainties associated with F_{max} with small sample sizes make it reasonable to conclude that the true significance level is within the .10 level for both the main effects. If the interaction is significant, it suggests that the relationship between performance and forced lockout is more complex than originally supposed.

Figure 4 provides another perspective on the relationships between the quality of solution and the effects of experience and lockout. It shows that the 5-min lockout and the variable lockout (averaging 5 min per trial) tended to raise the performance level of less-experienced subjects, while the Choice situation seemed to hinder them. Only the 8-min lockout tended to lower the performance level of more-experienced subjects.

If the indicated effects are real, perhaps a forced lockout serves as a facilitating *pacer* for less-experienced subjects. Experienced

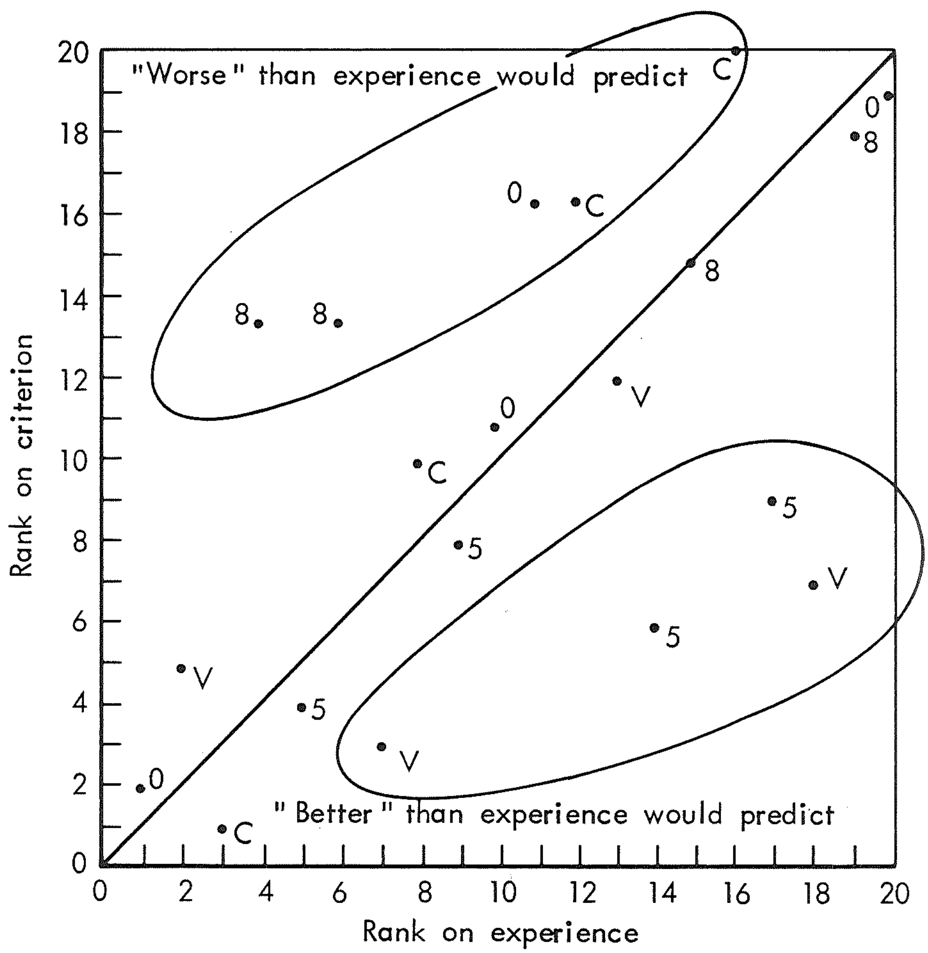


Fig. 4 —Rank on experience compared with rank on criterion measure

subjects appear to do well under any condition except the 8-min lockout, suggesting that an 8-min lockout may be disruptive, at least on this problem.

Turnaround or system response time was not a controlled factor in the experiment, being partly dependent on the number of people (in addition to our subjects) using JOSS. The differences being attributed to lockout times could have resulted from differences in system response speed.

An analysis of variance indicated that groups were not treated alike in terms of turnaround time; the difference was significant at the .01 level. However, Groups 0, 5, and 8 all have about equal turnaround times and their criterion scores cover the entire range. This suggests that response times did not significantly affect performance. A plot of scores versus average turnaround time indicated no obvious relationship. However, at the present time, it is not possible to assess accurately the implications of a relationship between response time and criterion score, if one exists, because not all differences in time are due to variances within the system. The subject may have specified more complex decision rules, which required more machine time to interpret, but which led to better results.

Variations in turnaround time durations (within subjects) were all very small and approximately equal, with the exception of those of three subjects. Of the three subjects with significantly larger variations, the one with the largest obtained the highest criterion score.

The Gold hypothesis is supported by the evidence presented thus far: lockout for a brief time tends to improve performance, at least for less-experienced subjects. Our corollary conjecture is also supported: a longer lockout period tends to degrade performance, at least for more-experienced subjects.

Time Factors

Whether or not quality of solutions differs beyond a chance level from group to group, it is of interest to compare the groups in terms of how much time was spent by man and machine in reaching the obtained solutions.

<u>Machine Time</u>. Table 3 summarizes the total number of trials run by each subject; Table 4 shows the results of an analysis of variance of these data. Lockout is significant at the .001 level, and the interaction between experience and lockout is significant at the .01 level. The main effect of experience is not significant.

Table 3

NUMBER OF TRIALS RUN BY SUBJECTS IN EACH GROUP

Group	0	5	8	V	C	Total
Higher Experience	15	7	8	8	14	--
	19	10	8	9	14	--
Σ	34	17	16	17	28	112
Lower Experience	20	5	8	10	7	--
	25	6	3	12	3	--
Σ	45	11	11	22	10	99
$\Sigma\Sigma$	79	28	27	39	38	211

Table 4

ANALYSIS OF VARIANCE SUMMARY: NUMBER OF TRIALS

Source	Sum of Squares	Degrees of Freedom	Mean Square	F	Significance Level
Total	634.95	19	--	--	--
Experience	8.45	1	8.45	1.74	NS
Lockout	452.45	4	113.11	23.32	.001
E × L	125.55	4	31.64	6.52	.010
Error	48.50	10	4.85	--	--

Inspection of the group totals in Table 3 indicates that the forced lockout groups all performed far fewer trials than the control group, with Groups 5 and 8 being the lowest and about equal. (These are the groups that scored highest and lowest, respectively, on the criterion measure.) Group 5 used approximately one-third the number of trials that the control group used. The effect of self-imposed constraint (Group C) is of lesser magnitude, but still appreciable. As suggested by the significant interaction, the lockout factor appears to affect the less-experienced subjects differently than it does the more-experienced ones.

It is important to recognize that it was physically possible for the subjects in the forced-lockout groups to perform many more trials than most of them attempted. The constraint appears to be at least partly psychological, as would be predicted from Carbonell's model--i.e., for these subjects, the "cost" of a computation was increased.

Had the calculation times been constant for every trial and for every subject, "number of trials" would be equivalent to "amount of computer time used in assessing inputs." Figure 5 shows the relationship between this theoretical computer time and the criterion score for each group, using average measures in both cases. This relationship provides a measure of problemsolving efficiency, at least for machine time. The superiority of Groups 5 and V is evident.

Another perspective on effectiveness of machine-time expenditure appears in Fig. 6. The plotted points show how the average best criterion score (based on each subject's "best score so far") for each group varied over the first ten trials.* By the end of the second trial, all experimental groups are clearly differentiated from the control group (Group 0); by the end of the fourth trial, major differences appear among the various experimental groups.

Console use can be measured approximately from the number of JOSS lines input and output by each subject. Figure 7 shows the input, output, and combined input and output totals and averages per trial for each of the five groups. The total number of lines input by the three lockout groups and the total number of lines output by two of them (Groups 5 and V) are less than the corresponding measures for the free-use and self-restraint groups. However, the differences here are not nearly so great as the differences in the number of trials run by the

---

*These curves do not represent a linear time progression, because intertrial times varied widely. Figure 8 shows the averages plotted against time.

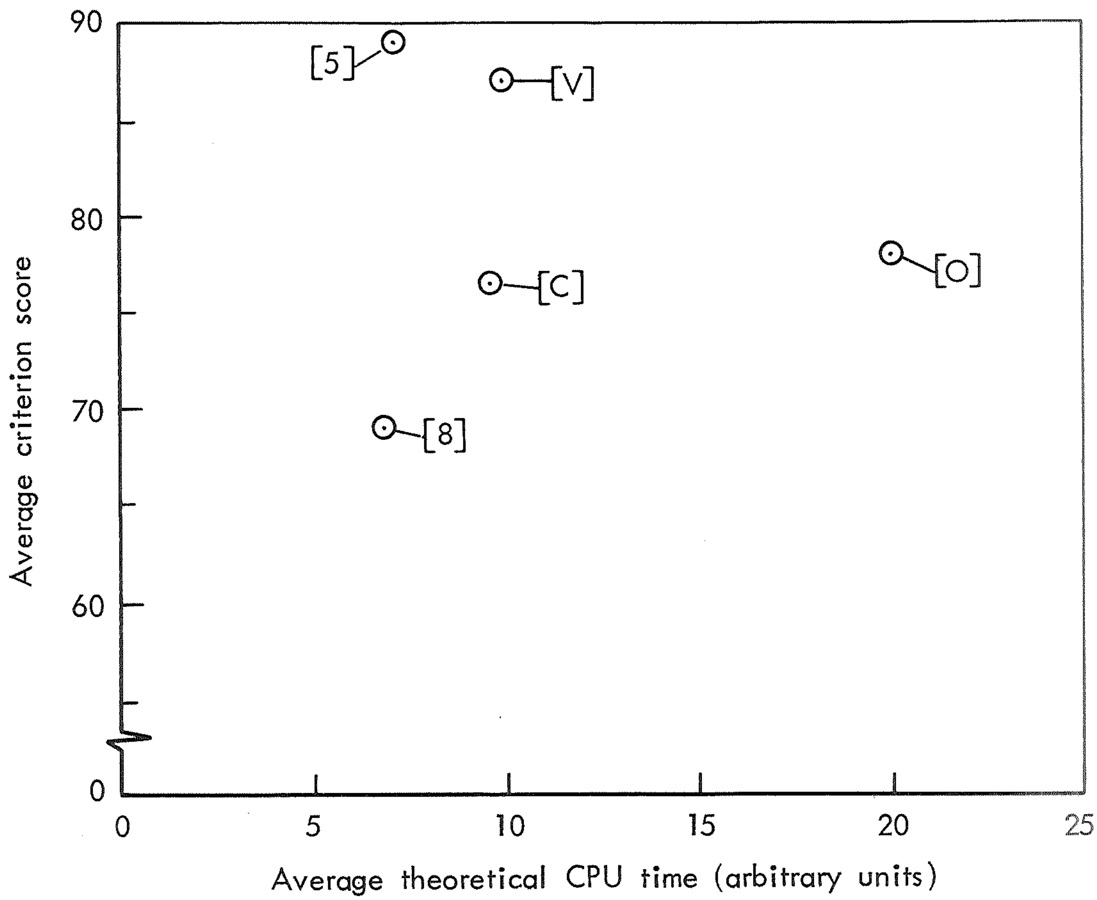


Fig.5 — Comparison by groups of average theoretical CPU time with average criterion score

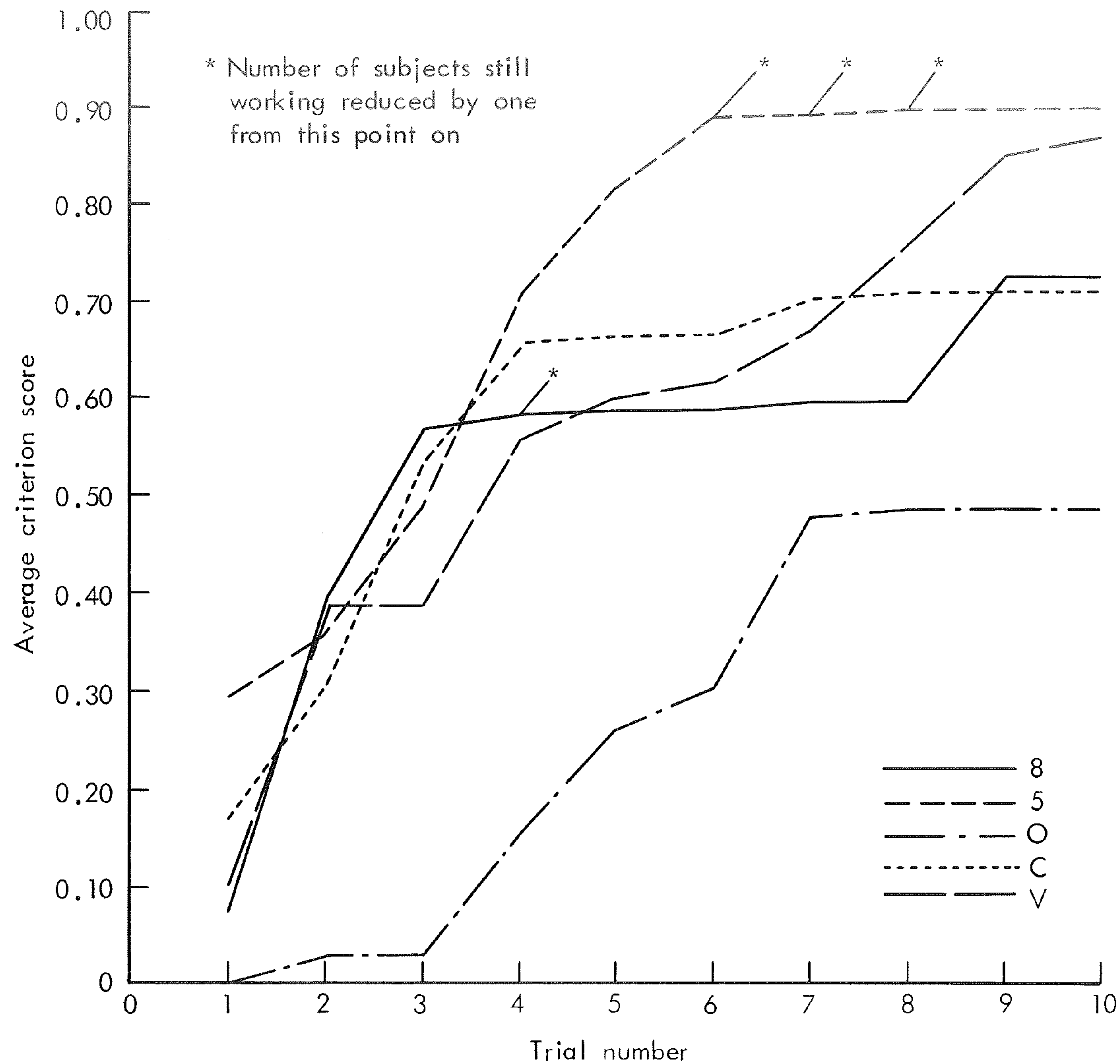


Fig. 6—Average of best scores for first ten trials

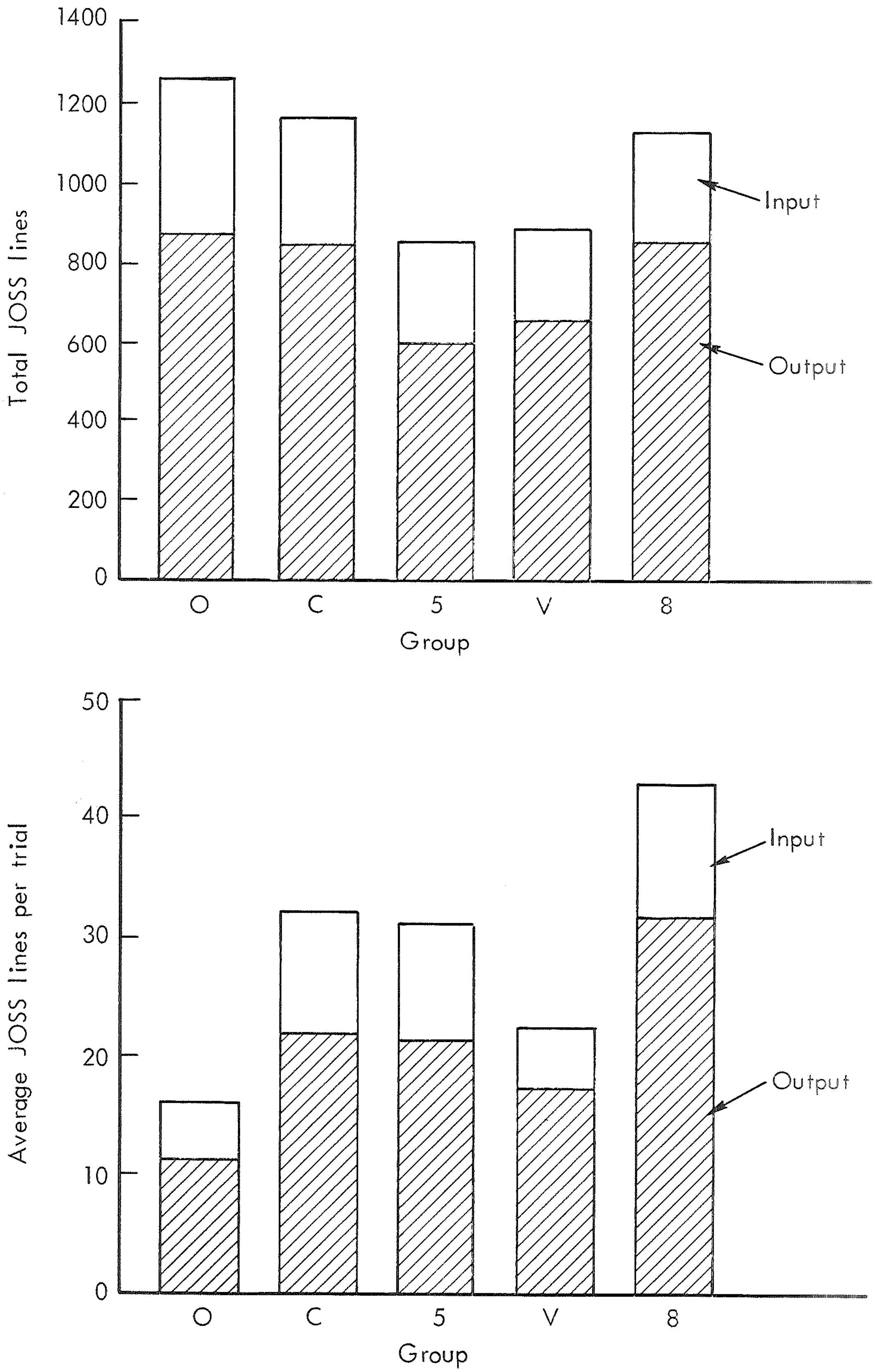


Fig.7 — Summary of terminal activity

groups. As indicated in the lower graph of Fig. 7, showing average input and output per trial, the lockout affected the "packaging" of communications more than it did the total volume. The free-use group tended to execute a new evaluation computation after making limited changes; the other groups did so only after reformulating more extensively.

User Time. One method of comparing the working efficiency of the several groups is to measure the total time each subject took to obtain his best solution. Although no performance criterion was set, the measure of time taken is meaningful when used in conjunction with the previously discussed measures--quality of solution and machine time.

Table 5 presents the times taken by each subject to achieve his best solution. An analysis of variance of these data showed lockout to be significant at the .05 level; neither experience nor the interaction between lockout and experience was significant. Group 5 subjects reached their best scores fastest; the other groups were slower and not too different from each other, the consecutive order being 8, V, C, and 0. This result is especially interesting because the Group 5 average score was the highest; these subjects not only got better scores, but they also got them faster and used less machine time in the process. The relatively poorer performance of the Group V subjects supports the contention that fixed intervals are less disruptive than variable ones.

Table 5

TIME TAKEN TO ACHIEVE BEST SOLUTION

Group	0	5	8	V	C	Total
Higher Experience	121	91	125	98	116	--
	122	83	78	106	125	--
Σ	243	174	203	204	241	1065
Lower Experience	111	36	105	107	113	--
	91	78	112	116	101	--
Σ	202	114	217	223	214	949
ΣΣ	445	288	420	427	455	2014

Figure 8 shows the average of best scores achieved by each group in 10-min intervals. These curves show the general improvement in performance usually associated with a learning situation, and also show some of the major differences in the efficiency of the manpower expenditures of the five groups. Had the good scores of Group 5 come early in the problem sessions, only to be followed by a long series of less effective solutions, one might suspect that random good fortune had played a major role in their success. However, with but one exception, the best trials came at or near the end of the series.

All four members of Group 5 elected to terminate their problem sessions after 95-100 min. They worked 20-30 min less than did the other subjects. Thus, they not only spent less time obtaining their best solutions, but also spent less time working on the problem. The behavior of Group 5 can be interpreted as supporting the model presented by Carbonell [5] in which the "decision to quit" is a function of both the cost of continuing and satisfaction with present results. A similar interpretation might account for the continuation of Groups 0 and C (the costs were lower) and Group 8 (the satisfaction with results was less).

Figure 9 presents the data from Table 5 in another perspective. Assuming some compensatory scale that permits trade-offs between quality of solution and time used to obtain that solution, it is possible to derive another measure of problemsolving efficiency. In Fig. 9, each subject's best score is plotted against the time he took to get that score. Diagonal line A indicates a 1:1 trade-off reference line; line B represents a .75:1 trade-off reference. All members of Group 5 are on the "more-efficient" side of both lines. The less stringent reference also includes three members of Group V, the top members of Groups C and O, and the second member of Group 8. Again, the data suggest that there is some efficiency in manpower, as well as machine time, resulting from the short lockout period.

Figures 8 and 9 also compare the results of a known, fixed lockout time with variable and unknown times of equal average duration. The data appear to favor the known, fixed interval slightly.

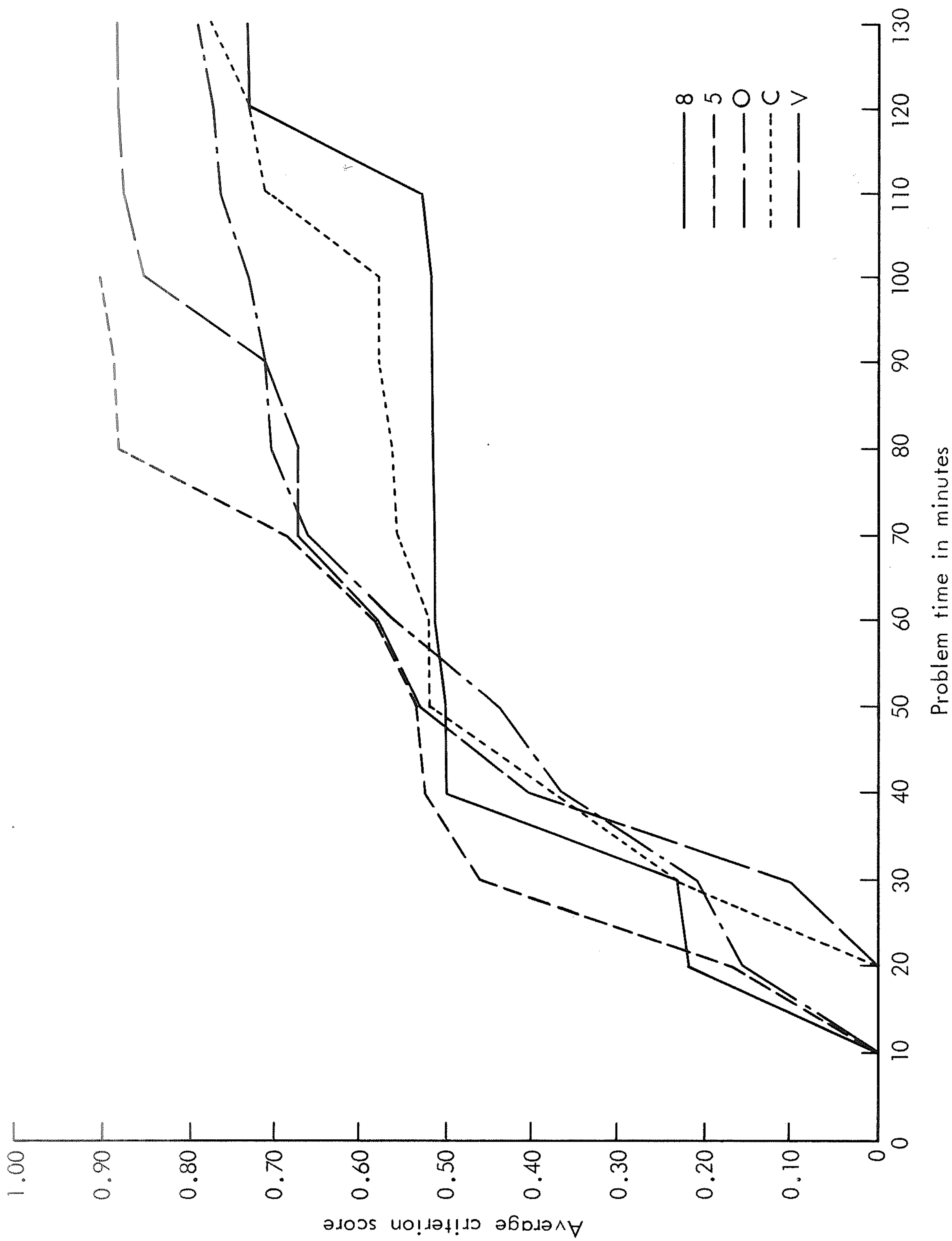


Fig. 8—Average of best scores over time

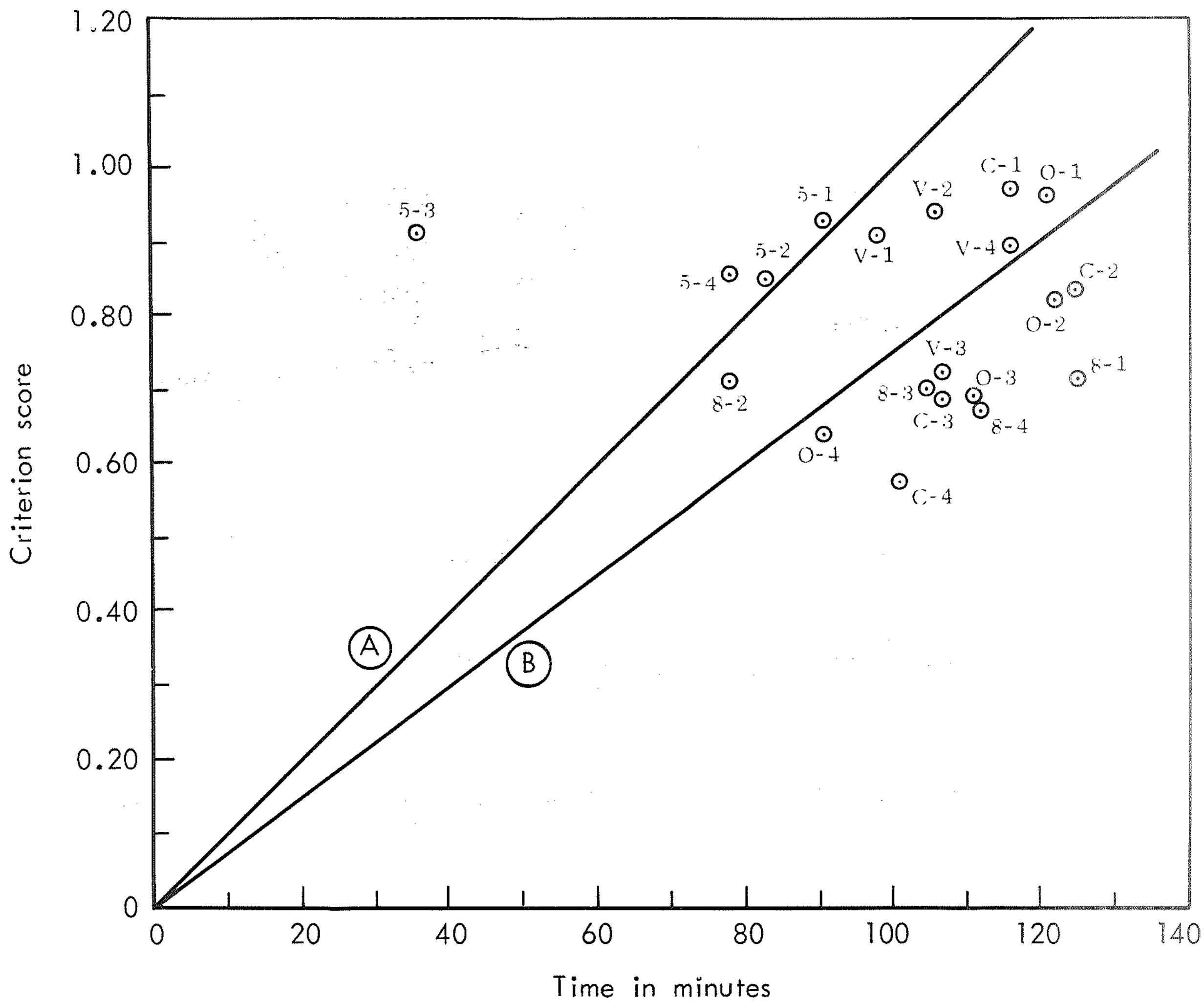


Fig.9 — Time to achieve best criterion score versus best criterion score

Interaction Patterns

Figure 10 shows the cumulative number of trials for each of the five experimental groups. The forced-lockout conditions and the self-imposed constraint all resulted in much lower interaction rates; the computer utilization of the control group (Group 0) is two to three times greater than that of the other groups.

Figure 11 shows the cumulative number of lines input to JOSS for each of the five groups. The rates appear to be fairly constant, with the slopes being less for the groups with forced lockout. As indicated earlier, the volume of communication did not vary as much as did the number of evaluation calculations (trials).

Three subjects (the two less-experienced members of Group C and the least experienced subject in Group 8) attempted to write complex decision rules. As a result, they worked a long time before attempting their first trial and performed few total trials. Such behavior could be interpreted in terms of the Carbonell [5] model as indicating "excessive cost" being placed on a computer trial. Other data (from the observer's report) suggest that this is precisely the problem faced by the Group 8 subject mentioned above. He would begin a trial, then stop, commenting that it was not a good enough hunch on which to "use up" one of his limited number of trials. In effect, he locked himself out of the system while searching for a better approach.

Figure 12 compares the post-trial behavior of all five groups. It shows the cumulative percentage of each group's trials that were followed by keyboard activity within a given number of minutes. A subject might or might not use the keyboard as soon as it came under his control. Often, however, a subject immediately using the system would merely note a single datum, then turn to other work for an extended period of time before again interacting with the system. Subjects with free access tended to do this more frequently than did those subject to lockout; the latter substituted pencil and paper for this "memo pad" function.

Figure 12 reveals that members of Group 0 (free use) and Group C (conservative use) often would have been prevented from interacting with

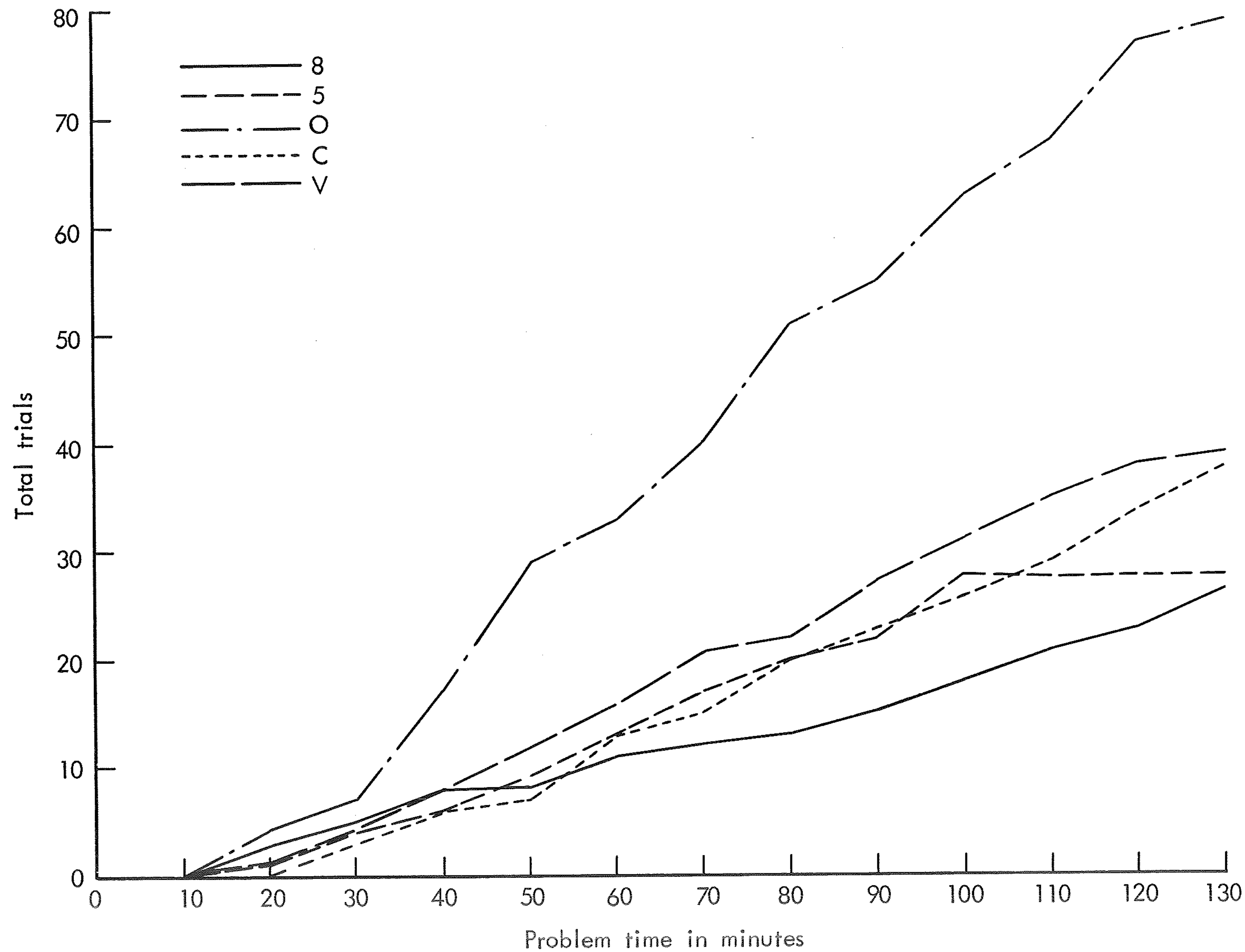


Fig. 10—Cumulative trials over time

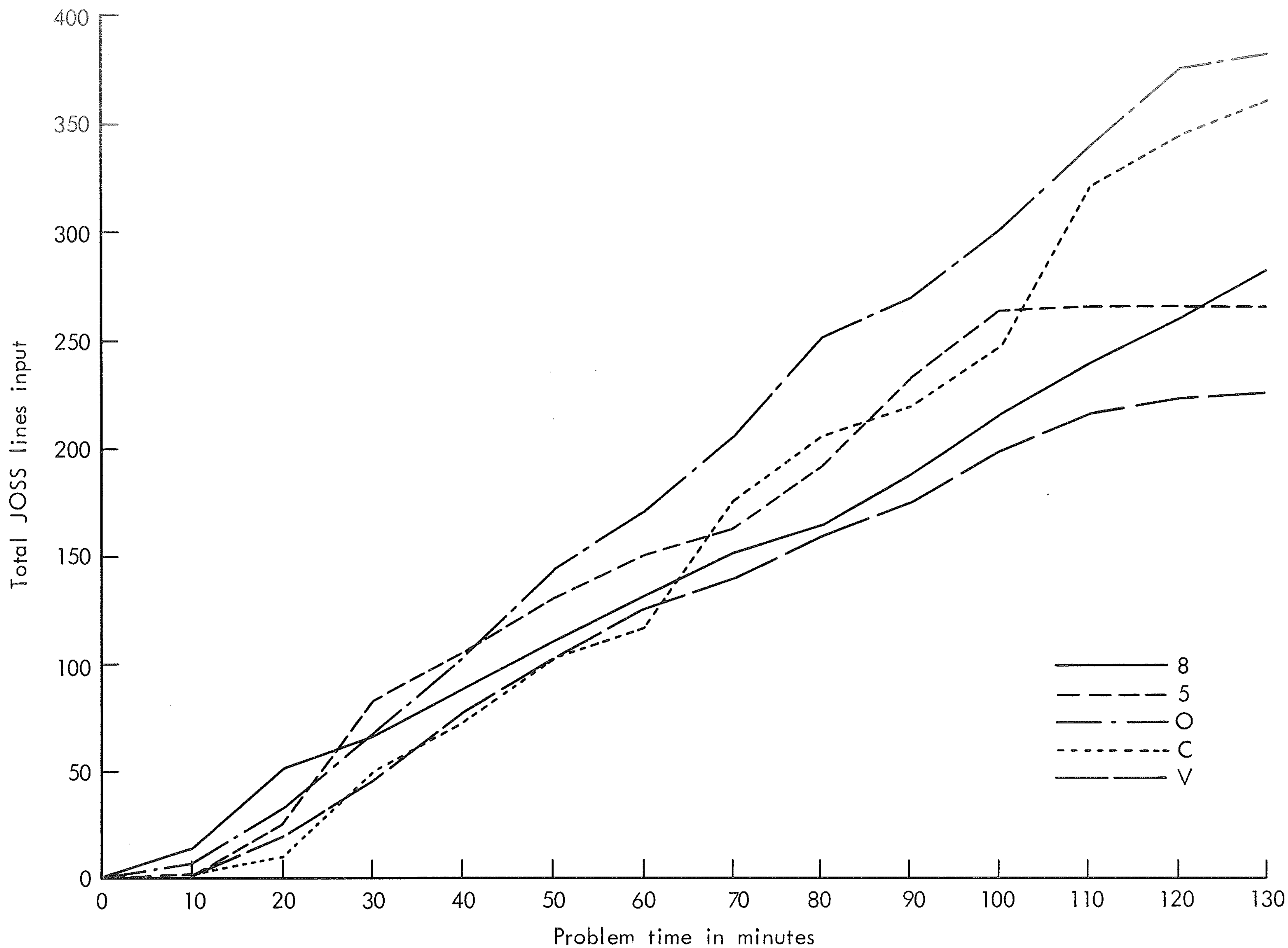


Fig. 11 — Cumulative JOSS input over time

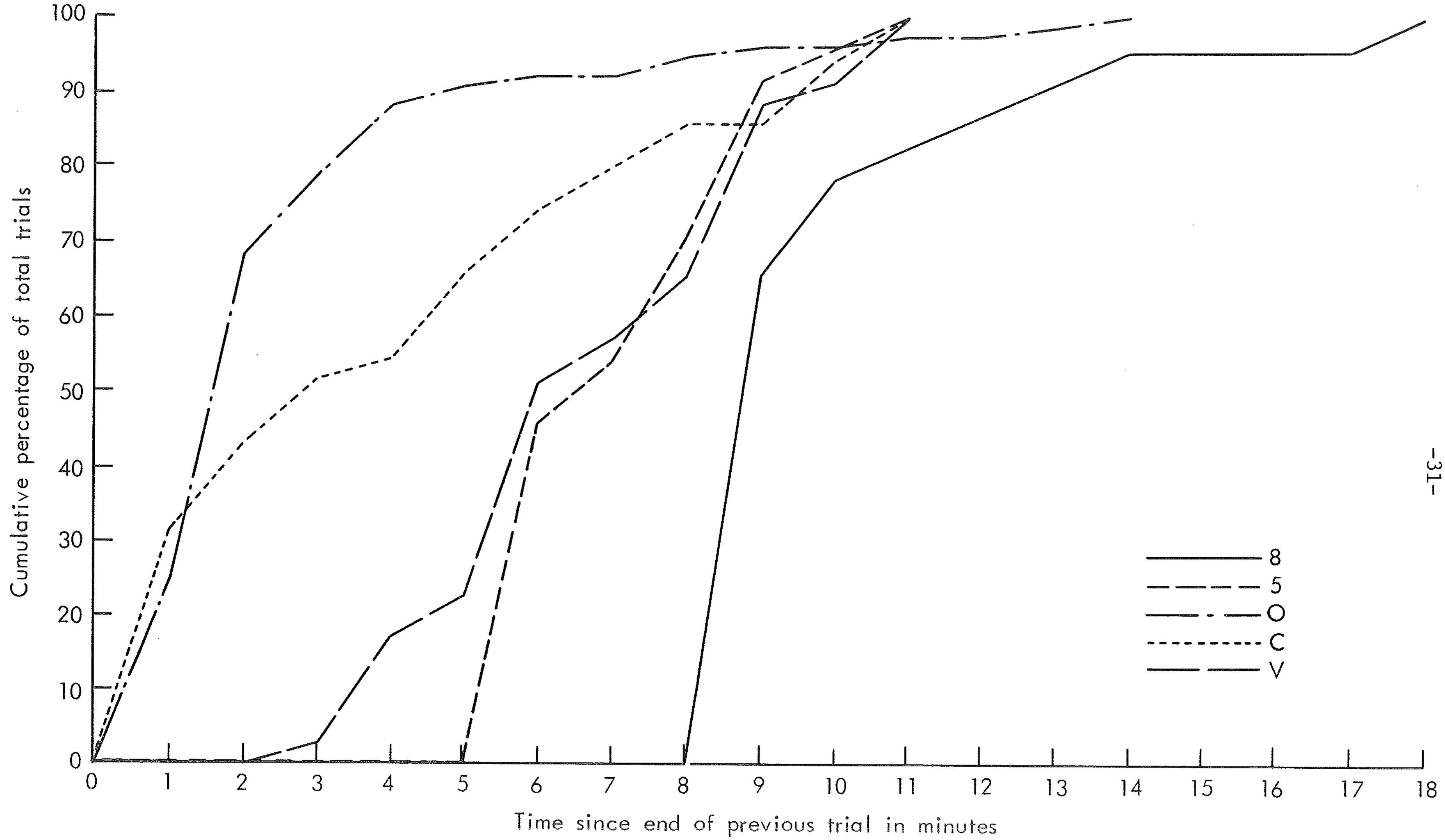


Fig. 12—Cumulative percentage of trials followed by keyboard activity within a given time

the system had they been subject to forced lockout. Comparison of the graphs for Groups 0 and C shows that the effect of the charging algorithm influenced not only number of trials run, but general keyboard activity as well.

If a subject in a lockout group began typing immediately following the end of the lockout, it might be inferred that the lockout was interferring physically with his problemsolving operations. On the other hand, if he did not begin typing for some minutes after the end of the lockout, one might infer that the physical effect of the lockout was not as significant. However, the absence of psychological pressures must not be presumed; the subject could still be operating in a manner other than his preferred style.

The differences in the intertrial intervals for the five groups can be seen in Fig. 13. The graphs show the cumulative percentage of total trials that occurred within the indicated number of minutes after the preceding trial. Eighty percent of the intertrial intervals were within 6 min for Group 0, 13 min for Group C, 14.5 min for Group V, 16 min for Group 5, and 18 min for Group 8. Times so different from that of the control group suggest that something more than delayed physical access was involved.

Summary of Experimental Findings

The results of the experimental portion of the study suggest that short periods of forced lockout between successive trials can both increase problemsolving effectiveness and reduce the amount of computer time and user time expended. Longer periods of lockout may reduce the effectiveness of more-experienced users. With experienced users, a self-imposed constraint, such as that resulting from a restrictive charging algorithm, can decrease machine time without reducing problemsolving effectiveness. Comparisons of fixed versus variable lockout intervals of similar average duration tend to favor the fixed-interval conditions.

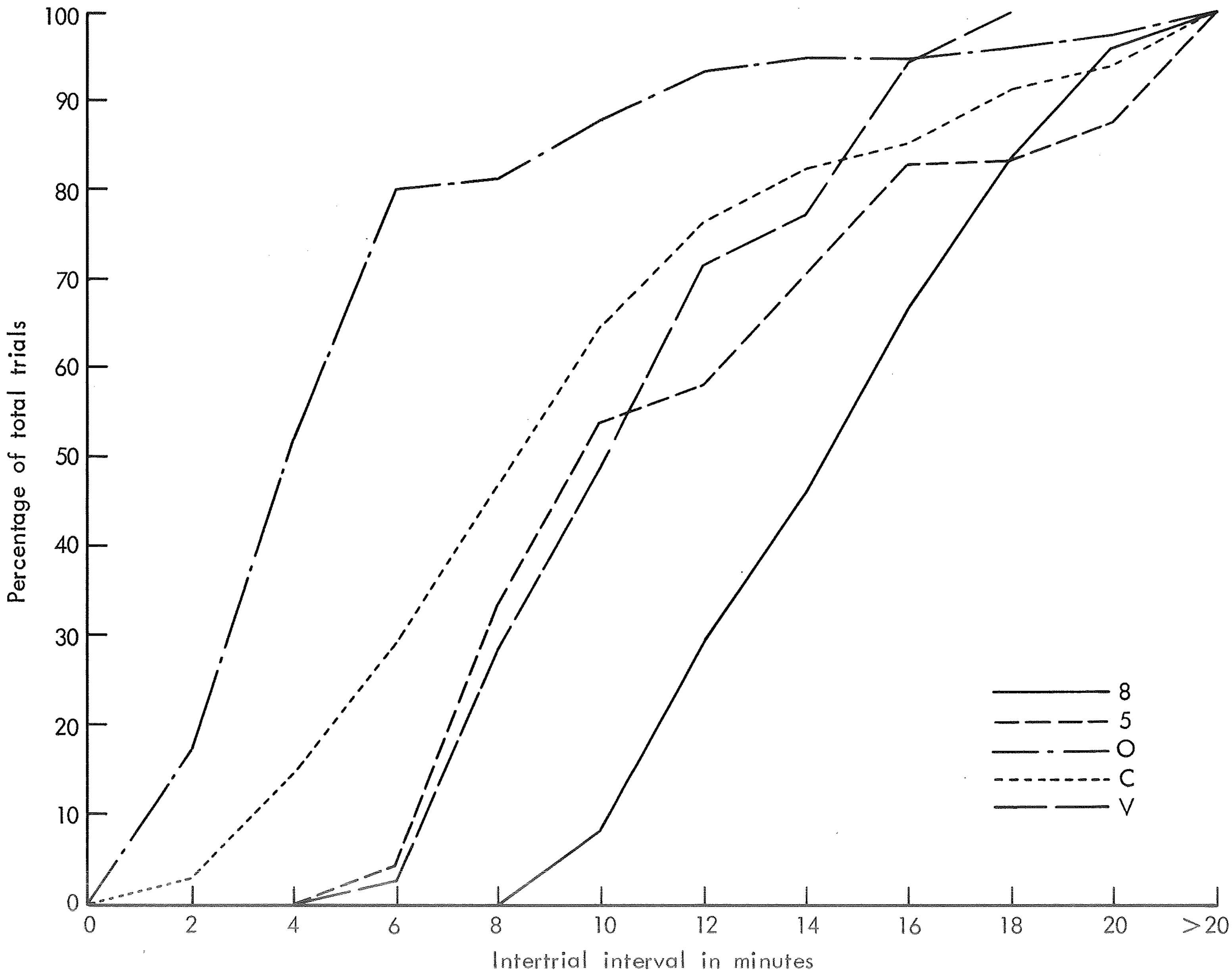


Fig. 13—Cumulative percentage of trials versus intertrial intervals

EXPLORATORY ANALYSES

To identify factors to be included in the design of subsequent experiments, several other exploratory analyses were carried out using the obtained performance measures. Table 6 presents the results of a stepwise multiple regression analysis (see Dixon [10]) to predict the criterion score (value of best solution) from 32 other measures.* The detailed regression information is presented for the step using seven predictor variables that accounted for 90 percent of the total variance of the criterion scores.

The variable labels cited in the summary table portion of Table 6 have the following identifications:

- EXPER: Experience score of the subject on an arbitrary scale from 60 to 98, with 98 being the highest.
- M/NONM: The ratio of time spent in machine-related activities to time spent in nonmachine-related activities.
- ANNOTE: The relative amount of annotation done by the subject on printouts and maps.
- Z CHGS: The number of trials on which decision rules were changed.
- IORAT: The ratio of JOSS lines input to JOSS lines output.
- 12+13: Number of maps and sheets of scratch paper used.
- TOTJOS: The total number of JOSS lines typed.
- XYTRYS: The total number of hospital locations tried.

---

*Other measures included time to obtain best score, trials to best score, total trials, JOSS lines in, JOSS lines out, time delay to first trial, number of matrices printed, number of maps used, number of scratch sheets used, number of hospital locations plotted on maps, number of trials on which hospital locations were changed, total number of input errors made, intertrial intervals (RMS), lockout intervals (0, 5, or 8), time used for machine-related work, and total time used. The remaining measures were functions or combinations of the measures noted.

Table 6

MULTIPLE REGRESSION ANALYSIS PREDICTING
SCORE ON CRITERION (N=20)

Step Number: 9
Variable Entered: 23
Multiple R: 0.9655
Standard Error of Estimate: 0.0396

Analysis of Variance

Source of Variation	Degrees of Freedom	Sum of Squares	Mean Square	F Ratio
Regression	7	0.259	0.037	23.557
Residual	12	0.019	0.002	--

Variables in Equation

Variable	Coefficient	Standard Error	F to Remove
(CONSTANT 0.61591)	--	--	--
TOTJOS 8	0.00027	0.00016	2.9013 (2)
IORAT 11	0.20164	0.04321	21.7758 (2)
EXPER 4	0.00445	0.00093	22.8610 (2)
Z CHGS 19	-0.03040	0.00477	40.6478 (2)
XYTRYS 23	0.00311	0.00192	2.6173 (2)
12+13 17	-0.01317	0.00457	8.3205 (2)
M/NONM 30	-0.36835	0.05177	50.6170 (2)

Summary Table

Step Number	Number Variables Used	Variable Entered	Variable Removed	Multiple Regression Coefficient (R)	R Squared (Corrected)	F to Enter or Remove
1	1	EXPER 4		0.5380	0.2894	7.3314
2	2	M/NONM 30		0.6530	0.3943	4.0590
3	3	ANNOTE 14		0.7808	0.5635	7.5112
4	4	Z CHGS 19		0.8480	0.6663	5.8483
5	5	IORAT 11		0.8994	0.7579	6.5733
6	6	12+13 17		0.9506	0.8692	12.7761
7	5		ANNOTE 14	0.9505	0.8776	0.0298
8	6	TOTJOS 8		0.9578	0.8879	2.2000
9	7	XYTRYS 23		0.9655	0.9009	2.6173

Of the relationships indicated, some may be trivial or spurious, and, of course, correlation does not imply causation. Any attempt to attach operational significance to these results would be pure speculation without subsequent experimental testing. For example, if it were found that those who did well on the problem also scored low on factor X, it would not necessarily follow that reducing X would lead to better operations. In fact, the opposite may be true--i.e., more X may be needed, because it is useful to those having difficulty solving a problem.

The equation derived in the preceding analysis did not include the crude lockout measure or its squared value. Although their "F to enter" was relatively high from step 2 to step 6 in the analysis, neither was the "highest" variable, the one entered into the equation. The addition of variable 17 (correlated .69 and .77 with the lockout measures) in step 6 greatly reduced their probability for subsequent selection.*

SUBJECTS' EVALUATIONS

Additional data were gleaned from the post-test questionnaires and the observer reports. The study revealed several weaknesses in the general method of post-test reconstruction by subjects of what they thought or what they did. Complaints made during the problem sessions were not necessarily repeated in the post-test question sessions. Approaches described in the post-test questionnaire as "initial approach" were more often approaches that had been adopted later, after one or more other approaches proved unsuccessful. Key insights that were verbalized and used in solving the problem were either not recalled or not remembered as important. Even the audio recordings were occasionally found to be at odds with reality; a subject might announce his intention to do one thing, then change his mind and do another without verbalizing the new intention.

Three of the items derived from the narrative data are summarized below.

---

*To obtain additional data on the relationships between the lockout variables and the other variables, the analysis was run again with the lockout variables forced into the regression equation. Their coefficients were reduced significantly by the successive addition of (a) total trials, (b) ratio of machine activity to nonmachine activity, (c) input errors, and (d) intertrial intervals.

Evaluation of the Problem

Fifteen subjects responded very positively to the problem; three others thought it was "fairly interesting"; and two were even less positive. The latter two were the subjects with the ninth best solution and the best solution.

Effect of Lockout

In response to a general question about system characteristics, nine of the twelve subjects who experienced forced lockout complained that it interfered with their problemsolving procedures. Two of the three remaining subjects had complained about the forced lockout orally while they were working on the problem. The last subject made no verbal complaints, but was observed attempting to type on a locked keyboard. Six of the subjects who complained about the lockout ranked in the top eight on the criterion measure; evidently satisfaction with the system was not prerequisite to success.

In addition to the general complaints, the subjects made some thought-provoking comments:

- "I'd do better if I did more trial and error, but I didn't think it was worth it." (Group 5)
- "Without lockout, I would try many more examples without too much thought." (Group V)
- "I used the time to plan the next move; it bothered me only at the end." (Group 5)
- "It helped my planning; I used the time more wisely. But it interfered with my operations, especially at the end." (Group 5)
- "It had one positive effect--it gave me time to check my work." (Group V)

Of particular interest is the implication that lockout *permits* a more careful, thoughtful approach. These comments and some of the operational data suggest that some people feel compelled to interact with a system if it is available.

Subjective Time Estimates

A comparison of data from the observer reports and JOSS records with estimates given by the subjects in the post-test questionnaires indicated that, for some subjects, perceived time differed greatly from actual time spent in various activities. For example, in estimating time spent waiting for computer output, seven of the twenty subjects overestimated the time and three underestimated it by 50 percent or more. Of those overestimating, one was off by a factor of 10, one by a factor of 4, and three by a factor of 2. Of those underestimating, one was off by a factor of 2.5 and two were off by a factor of 2. There was no relationship between error of estimate and group membership or criterion score.

IV. IMPLICATIONS FOR FUTURE WORK

The statement of experimental hypotheses or behavioral theories bearing on the man-machine-problem interface is hindered by the absence of a common rubric for discussing the three subareas. Work on human problemsolving, creativity, and decisionmaking tends to employ terms and concepts derived from early learning theory. Studies or descriptions of machine capabilities and characteristics are couched in terms specific to a particular machine context. Schema for analyzing problems--for specifying the characteristics that differentiate problems--are both rare and lacking in power. Until problems, problemsolvers, and problemsolving devices can be described in a common reference system, insights on the nature of their interactions will remain scarce.

NORMATIVE ASPECTS

The method used in the present study to assess subjects' experience/capability appeared to be valid and adequate for general studies of system effectiveness, but it lacked the objectivity needed for replication and the level of detail needed for analysis. A finer-grain measurement of both capabilities and performance might uncover some important relationships.

The results of the study, together with observations made during and after the problem sessions, suggest that the normative data should be extended to include measures of personality factors and attitudes, in addition to problemsolving capability. A sufficiently broad data base could yield results bearing on selection and training of personnel, as well as on computer system design.

METHODOLOGICAL ASPECTS

To obtain empirical statistics on leading parameters of user performance, it is necessary first to identify those parameters. In searching for a more effective model of the problemsolving situation, it may be necessary to forego some of the luxuries of "neat" behavioral experimentation. Narrative data from trained or partially trained subjects could provide important clues for guiding future gathering of objective data.

In performing the study described here, we experienced many of the difficulties inherent in attempting to secure real-time introspective data on the problemsolving process and in having an observer interact, within bounds, with the subjects. Prior training of the subjects could minimize the contamination of these fruitful sources of data.

Another source of difficulty in this study was lack of control over the machine system. Dedicating the entire system to the study would increase experimental rigor, but a more practical alternative may be available. Similarly, greater advantage could be taken of the computer system's capability for automatically generating and processing data compilations of interest. (For example, it would be easy to evaluate every decision rule and every set of hospital locations relative to optimum standards.)

The present study suffered from necessary limitations in size and scope. Insights are generated by concatenations of events; it is hoped that our future studies in this domain will include more than one type of problem to be solved, more than one type of terminal equipment, more than one type of work environment, and a broader range of time variables. (After seeing the data, we lamented the elimination of the originally planned 1-min and 2-min lockouts.)

For the present, one can only speculate about the interrelationships between different types of problems and different types of terminal equipment, about the advantages and disadvantages of public and isolated work areas, and whether or not lockout times close to the average unrestricted response times would produce the beneficial effects associated with the 5-min interval. It is possible that even a very short interruption of noticeable duration would promote more productive and efficient man-machine interaction.

V. CONCLUDING REMARKS

The results of this exploratory study raise some interesting questions regarding popular beliefs about the domain of man-machine problemsolving. The evidence suggests that, in this experimental context at least, users are dissatisfied by a mild restraint on their free interaction with the computer. However, they also tend to solve problems more effectively, using less computer time and less of their own time in the process. One might begin to suspect that shibboleths such as "faster is better" and "more computer time means less human time" may serve the computer salesman better than the consumer. The results also challenge the validity of user acceptance as a general index of system effectiveness. The user may want what inconveniences him least in the short run, or he may want something he has been led to believe that he *should* want, but the general efficacy of such desires is unclear.

Definitive answers to questions regarding the nature of relevant parameters of problemsolving systems are of more than academic value. For example, under some circumstances, an organization under pressure to expand its hardware inventory to meet increased demand might find it far more productive to keep its current system and introduce some form of constraint (e.g., an accounting or control system) that will encourage more judicious and creative use of the existing computational capabilities. However, without more information and a more complete understanding, it would be a mistake to designate either approach as the "right" one. The present work clearly demonstrates that the relationships involved in man-machine problemsolving are neither obvious nor simple and that further investigation could have practical significance.

REFERENCES

1. Sackman, H., *Experimental Investigation of User Performance in Time-Shared Computing Systems: Retrospect, Prospect, and the Public Interest,* System Development Corporation, Santa Monica, Calif., SP-2846, 5 May 1967.
2. Sackman, H., *Man-Computer Problem Solving,* Auerbach Publishers, New York, 1970.
3. Gold, M. M., "Time-Sharing and Batch-Processing: An Experimental Comparison of Their Values in a Problem-Solving Situation," *Comm. ACM,* Vol. 12, No. 5, May 1969, pp. 249-259.
4. Gold, M. M., "Methodology for Evaluating Time-Shared Computer Usage," Ph.D. Thesis, Alfred P. Sloan School of Management, Massachusetts Institute of Technology, 1967.
5. Carbonell, J. R., "On Man-Computer Interaction: A Model and Some Related Issues," *IEEE Trans. Systems Science and Cybernetics,* Vol. SSC-5, No. 1, January 1969, pp. 16-26.
6. Carbonell, J. R., J. I. Elkind, and R. S. Nickerson, "On the Psychological Importance of Time in a Time Sharing System," *Human Factors,* Vol. 10, No. 2, April 1968, pp. 135-142.
7. Sackman, H., and M. M. Gold, *Time-Sharing versus Batch-Processing: An Experimental Inquiry into Human Problem-Solving,* System Development Corporation, Santa Monica, Calif., SP-3110, 1968.
8. Nickerson, R. S., J. I. Elkind, and J. R. Carbonell, "Human Factors and the Design of Time Sharing Computer Systems," *Human Factors,* Vol. 10, No. 2, April 1968, pp. 127-134.
9. Myers, J. L., *Fundamentals of Experimental Design,* Allyn and Bacon, Boston, 1966.
10. Dixon, W. J. (ed.), *BMD Biomedical Computer Programs,* University of California Press, Berkeley and Los Angeles, 1968.